Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands*

Josef Dick
University of New South Wales

July 9, 2010

Abstract

We study numerical approximations of integrals $\int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ by averaging the function at some sampling points. Monte Carlo (MC) sampling yields a convergence of the root mean square error (RMSE) of order $N^{-1/2}$ (where N is the number of samples). Quasi-Monte Carlo (QMC) sampling on the other hand achieves a convergence of order $N^{-1+\varepsilon}$, for any $\varepsilon > 0$. Randomized QMC (RQMC), a combination of MC and QMC, achieves a RMSE of order $N^{-3/2+\varepsilon}$. A combination of RQMC with local antithetic sampling achieves a convergence of the RMSE of order $N^{-3/2-1/s+\varepsilon}$ (where $s \geq 1$ is the dimension). QMC, RQMC and RQMC with local antithetic sampling require that the integrand has some smoothness (for instance, bounded variation). Stronger smoothness assumptions on the integrand do not improve the convergence of the above algorithms further.

This paper introduces a new RQMC algorithm, for which we prove that it achieves a convergence of the RMSE of order $N^{-\alpha-1/2+\varepsilon}$ if the integrand has square integrable partial mixed derivatives up to order α in each variable. Known lower bounds show that this rate of convergence cannot be improved. We provide numerical examples for which the RMSE converges approximately with order $N^{-5/2}$ and $N^{-7/2}$, in accordance with the theoretical upper bound.

*MSC 2000: Primary: 65C05; Secondary: 65D32; Keywords: Digital nets; randomized quasi-Monte Carlo; quasi-Monte Carlo; Acknowledgement: The author is supported by an Australian Research Council Queen Elizabeth II Fellowship.

1 Introduction

The problem of numerically integrating a function comes up frequently in science and engineering. We consider the standardized problem of approximating the integral $\int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, that is, we assume that any transformations necessary to change from different domains and density functions have already been carried out. Monte Carlo algorithms use i.i.d. uniformly distributed samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in [0,1]^s$ to approximate the integral by $\frac{1}{N}\sum_{n=1}^N f(\boldsymbol{x}_n)$. For functions $f \in L_2([0,1]^s)$ the Monte Carlo method has a root mean square error (RMSE) of $\mathcal{O}(N^{-1/2})$. An alternative to Monte Carlo is Quasi-Monte Carlo. In this method one designs sample points which are more uniformly distribution with respect to some criterion (in one dimension this criterion is the Kolmogorov Smirnov distance between the uniform distribution and the sample point distribution). These achieve a worst case error which decays with $\mathcal{O}(N^{-1+\varepsilon})$ for any $\varepsilon > 0$, see [7]. Owen [15, 16, 17] introduced a randomization of QMC which achieves a RMSE of $\mathcal{O}(N^{-3/2+\varepsilon})$. Owen's randomization method uses a permutation applied to digital nets (which is a construction scheme for sample points used in quasi-Monte Carlo) called scrambling. A slight improvement of Owen's scrambling method of digital nets can be obtained by combining this approach with local antithetic sampling, see [19]. Therein it was shown that one obtains a convergence of the RMSE of $\mathcal{O}(N^{-3/2-1/s+\varepsilon})$ (s is the dimension of the domain). The latter three methods require that the function f has some smoothness (for instance continuous partial mixed derivatives up to order 1 in each coordinate in the first two methods and continuous partial mixed derivatives up to order 2 in each coordinate in the third method). No further improvement on the rate of convergence is obtained when one assume that the integrand has continuous higher order partial mixed derivatives in each variable.

In this paper we introduce a randomization of quasi-Monte Carlo algorithms (which use digital nets as quadrature points) such that the RMSE converges with $\mathcal{O}(N^{-\alpha-1/2+\varepsilon})$ (for any $\varepsilon > 0$) if the integrand has square integrable partial mixed derivatives up to order α in each variable. This result holds for any $\alpha > 0$ and it is known that this result is best possible, see [14].

For the reader familiar with scrambled digital nets, we briefly describe the algorithm. The details on scrambled digital nets will be given in the next section.

1.1 The algorithm

The underlying idea of the new randomized QMC algorithm stems from [4, 5]. Central to this method is the digit interlacing function with interlacing factor $d \in \mathbb{N}$ given by

$$\mathscr{D}_d : [0,1)^d \quad \to \quad [0,1)$$

$$(x_1, \ldots, x_d) \quad \mapsto \quad \sum_{a=1}^{\infty} \sum_{r=1}^{d} \xi_{r,a} b^{-r-(a-1)d},$$

where $x_r = \xi_{r,1} b^{-1} + \xi_{r,2} b^{-2} + \cdots$ for $1 \le r \le d$. Let $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1} \in [0,1)^{ds}$ be a randomly scrambled digital (t, m, ds) net over the finite field $\mathbb{Z}_b$ of prime order b (we present the theoretical background on scrambled digital nets in the next section). Let $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,ds})$. Then one simply uses the sample points

$$\boldsymbol{y}_n = \big(\mathscr{D}_d(x_{n,1}, \ldots, x_{n,d}), \ldots, \mathscr{D}_d(x_{n,d(s-1)+1}, \ldots, x_{n,ds})\big) \in [0,1)^s \quad \text{for } 0 \le n < b^m.$$

The integral is then estimated using

$$\widehat{I}(f) = \frac{1}{b^m} \sum_{n=0}^{b^m-1} f(\boldsymbol{y}_n).$$

In Theorem 8 we show that if the integrand has square integrable partial mixed derivatives of order $\alpha \ge 1$ in each variable, then the variance of $\widehat{I}(f)$ satisfies

$$\mathrm{Var}[\widehat{I}(f)] = \mathcal{O}(N^{-2\min(d,\alpha)-1+\varepsilon})$$

for any $\varepsilon > 0$, where $N = b^m$ is the number of sample points.

Since scrambled digital nets (based on Sobol points) are included in the statistics toolbox of Matlab, this method is very easy to implement (an implementation can be found at http://quasirandomideas.wordpress.com.)

1.2 Numerical Results

Before we introduce the theoretical background, we present some simple numerical results which verify the convergence results.

Example 1

In this example the dimension is 1 and the integrand is given by $f(x) = x\mathrm{e}^x$. Figure 1.2 shows the RMSE from 300 independent replications. Here, the straight lines show the functions $N^{-3/2}$, $N^{-5/2}$ and $N^{-7/2}$. The other lines are the RMSE where the digit interlacing factor d is given by 1 for the upper dashed line, 2 for the dashed line in the middle and 3 for the lowest of the dashed lines. Figure 1.2 shows that in each case the RMSE converges approximately with order $N^{-d-1/2}$ (for large enough N). (The result for $d = 1$ appears to perform even better than $N^{-3/2}$.)

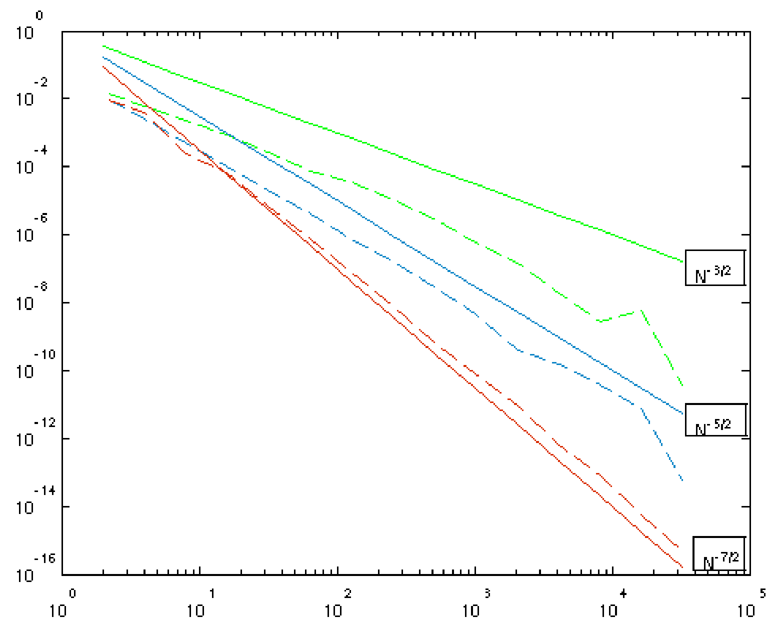

Figure 1: The green lines show $N^{-3/2}$ and the standard deviation where $d = 1$, the blue lines show $N^{-5/2}$ and the standard deviation where $d = 2$ and the red lines show $N^{-7/2}$ and the standard deviation where $d = 3$.

Example 2

We consider now a 2 dimensional example where the integrand is given by $f(x, y) = \frac{y\mathrm{e}^{xy}}{\mathrm{e}-2}$. This function was also used in [19] where the sample points are obtained by scrambling and local antithetic sampling.

Figure 1.2 shows again the RMSE for 300 independent replications. The straight lines show the functions $N^{-3/2}$ and $N^{-5/2}$. The two dashed lines show the RMSE when $d = 1$ (upper dashed line) and when $d = 2$ (lower dashed line). Figure 1.2 shows that in each case the RMSE converges approximately with order $N^{-d-1/2}$ (for large enough N).

In the following section we give the necessary background on QMC, digital nets, scrambling and Walsh functions. We then prove in Section 3 what can be observed from the numerical results, namely, that if the integrand has square integrable partial mixed derivatives of order α in each variable, then we obtain a convergence of the RMSE of $\mathcal{O}(N^{-\min(\alpha,d)-1/2+\varepsilon})$ for any $\varepsilon > 0$. A short discussion of the results is presented in Section 4. Some properties of the digit interlacing function $\mathscr{D}_d$ necessary for the proof is presented in Appendix A and a technical

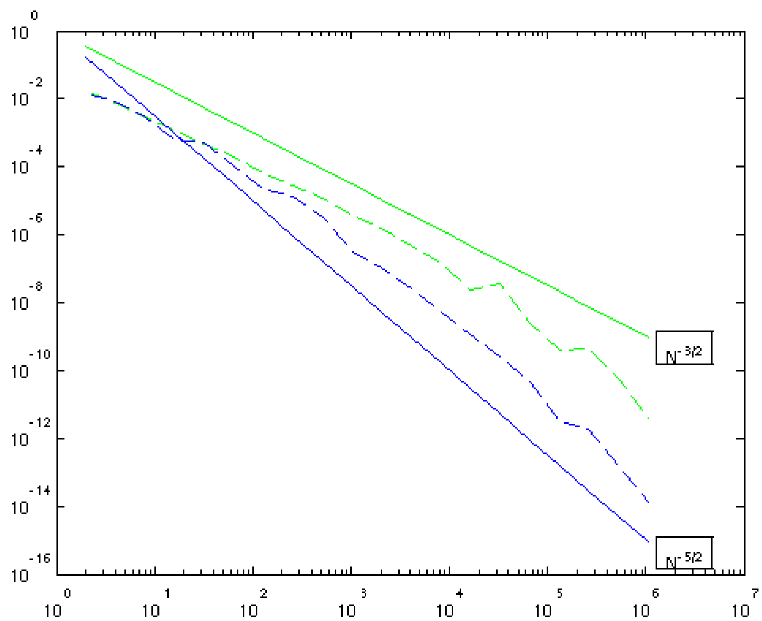

Figure 2: The green lines show $N^{-3/2}$ and the standard deviation where $d = 1$, the blue lines show $N^{-5/2}$ and the standard deviation where $d = 2$.

proof on the convergence of the Walsh coefficients is presented in Appendix B.

2 Background and notation

In this section we give the necessary background on QMC methods. Some notation is required, which we now present.

In this section, $c, C > 0$ stand for generic constants which may differ in different places.

Throughout the paper we assume that $b \geq 2$ is a prime number. We always have $\boldsymbol{k} = (k_1, \ldots, k_s)$, $\boldsymbol{k}' = (k'_1, \ldots, k'_s)$, $\boldsymbol{x} = (x_1, \ldots, x_s)$, $\boldsymbol{y} = (y_1, \ldots, y_s)$, $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,s})$, $\boldsymbol{y}_n = (y_{n,1}, \ldots, y_{n,s})$.

2.1 Quasi-Monte Carlo

Quasi-Monte Carlo algorithms $\widehat{I}(f) = \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{x}_n)$ are used to approximate integrals $I(f) = \int_{[0,1]^s} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$. The difference to Monte Carlo is the method by which the sample points

$\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1} \in [0,1)^s$ are chosen. The aim of QMC is to chose those points such that the integration error

$$\left| \int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{x}_n) \right|$$

achieves the (almost) optimal rate of convergence as $N \to \infty$ for a class of functions $f : [0,1]^s \to \mathbb{R}$. For instance, for the set of all such functions f which have bounded variation in the sense of Hardy and Krause, which we write as $\|f\|_{\mathrm{HK}} < \infty$, it is known that the best rate of convergence for the worst case error is

$$e = \sup_{f, \|f\|_{\mathrm{HK}} < \infty} \left| \int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{x}_n) \right| \asymp N^{-1+\varepsilon} \quad \text{for all } \varepsilon > 0.$$

(More precisely, there are constants $c, C > 0$ such that $cN^{-1}(logN)^{(s-1)/2} \leq e \leq CN^{-1}(\log N)^{s-1}$, see [7].)

Choosing the points $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1} \in [0,1)^s$ randomly as in MC, does not yield this rate of convergence. Even if a function has bounded variation in the sense of Hardy and Krause one obtains only a convergence of order $N^{-1/2}$ for randomly chosen sample points.

There is an explicit construction of the sample points $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}$ for which the optimal rate of convergence is achieved. The essential insight is that the quadrature points need to be more uniformly distributed than what one obtains by choosing the sample points by chance. One criterion for how uniformly a set of points $P_N = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}\}$ is distributed is the star discrepancy

$$D_N^*(P_N) = \sup_{\boldsymbol{z} \in [0,1]^s} \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_{\boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{z})} - \mathrm{Vol}([\boldsymbol{0}, \boldsymbol{z})) \right|,$$

where $[\boldsymbol{0}, \boldsymbol{z}) = \prod_{i=1}^s [0, z_i)$ with $\boldsymbol{z} = (z_1, \ldots, z_s)$, $\mathrm{Vol}([\boldsymbol{0}, \boldsymbol{z})) = \prod_{i=1}^s z_i$, the volume of $[\boldsymbol{0}, \boldsymbol{z})$ and

$$1_{\boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{z})} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{z}), \\ 0 & \text{otherwise.} \end{cases}$$

When $s = 1$ this becomes the Kolmogorov-Smirnov distance between the empirical distribution of the points and the uniform distribution. Further we call

$$\delta_{P_N}(\boldsymbol{z}) = \frac{1}{N} \sum_{n=0}^{N-1} 1_{\boldsymbol{x}_i \in [\boldsymbol{0}, \boldsymbol{z})} - \mathrm{Vol}([\boldsymbol{0}, \boldsymbol{z}))$$

the local discrepancy (of P_N).

The connection of this criterion to the integration error is given by the Koksma-Hlawka inequality

$$\left| \int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - \frac{1}{N} \sum_{n=0}^{N-1} f(\boldsymbol{x}_n) \right| \le D_N^*(P_N) \|f\|_{\mathrm{HK}}.$$

An explicit construction of point sets $P_N = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}\} \in [0,1)^s$ for which $D_N^*(P_N) \le CN^{-1}(\log N)^{s-1}$ is given by the concept of digital nets, which we introduce in the next subsection. Notice that for such a point set, the Koksma-Hlawka inequality implies the optimal rate of convergence of the integration error, since for a given integrand, the variation $\|f\|_{\mathrm{HK}}$ does not depend on P_N and N.

2.2 Digital nets

A comprehensive introduction to digital nets can be found in [7, 13].

The aim is to construct a point set $P_N = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}\}$ such that the star discrepancy satisfies $D_N^*(P_N) \le CN^{-1}(\log N)^{s-1}$. To do so, we discretize the problem by choosing the point set P_N such that the local discrepancy $\delta_{P_N}(\boldsymbol{z}) = 0$ for certain $\boldsymbol{z} \in [0,1]^s$ (those $\boldsymbol{z}$ in turn are chosen such that the star discrepancy of P_N is small, as we explain below).

It turns out that, when one chooses a base $b \ge 2$ and $N = b^m$, then for every natural number m there exist point sets $P_{b^m} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$ such that $\delta_{P_{b^m}}(\boldsymbol{z}) = 0$ for all $\boldsymbol{z} = (z_1, \ldots, z_s)$ of the form

$$z_i = \frac{a_i}{b^{d_i}} \quad \text{for } 1 \le i \le s,$$

where $0 < a_i \le b^{d_i}$ is an integer and $d_1 + \cdots + d_s \le m - t$ with $d_1, \ldots, d_s \ge 0$. Crucially, the value of t can be chosen independently of m (but depends on s). A point set P_N which satisfies this property is called a (t, m, s)-net in base b. An equivalent description of (t, m, s)-nets in base b is given in the following definition.

Definition 1 *Let $b \ge 2$, $m, s \ge 1$ and $t \ge 0$ be integers. A point set $P_{b^m} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\} \subset [0,1)^s$ is called a (t, m, s)-net in base b, if for all nonnegative integers $d_1, \ldots, d_s$ with $d_1 + \cdots + d_s \le m - t$, the elementary interval*

$$\prod_{i=1}^{s} \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right)$$

contains exactly b^{m-t} points of P_{b^m} for all integers $0 \le a_i < b^{d_i}$.

It can be shown that a (t, m, s)-net in base b satisfies

$$D_N^*(P_N) \leq C \frac{m^{s-1}}{b^{m-1}},$$

see [7, 13] for details.

Explicit constructions of (t, m, s)-nets can be obtained using the digital construction scheme. Such point sets are then called digital nets (or digital (t, m, s)-nets if the point set is a (t, m, s)-net).

To describe the digital construction scheme, let b be a prime number and let $\mathbb{Z}_b$ be the finite field of order b (a prime power and the finite field $\mathbb{F}_b$ could be used as well). Let $C_1, \ldots, C_s \in \mathbb{Z}_b^{dm \times m}$ be s matrices of size $dm \times m$ with elements in $\mathbb{Z}_b$ and $d \in \mathbb{N}$. The ith coordinate $x_{n,i}$ of the nth point $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,s})$ of the digital net is obtained in the following way. For $0 \leq n < b^m$ let $n = n_0 + n_1 b + \cdots + n_{m-1} b^{m-1}$ be the base b representation of n. Let $\vec{n} = (n_0, \ldots, n_{m-1})^\top \in \mathbb{Z}_b^m$ denote the vector of digits of n. Then let

$$\vec{y}_{n,i} = C_i \vec{n}.$$

For $\vec{y}_{n,i} = (y_{n,i,1}, \ldots, y_{n,i,dm})^\top \in \mathbb{Z}_b^{dm}$ we set

$$x_{n,i} = \frac{y_{n,i,1}}{b} + \cdots + \frac{y_{n,i,dm}}{b^{dm}}.$$

The construction described here is slightly more general to the classical concept to suit our needs (the classical construction scheme uses $d = 1$). In this framework we have that if $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$ is a digital (t, m, ds)-net, then $\{\mathscr{D}_d(\boldsymbol{x}_0), \ldots, \mathscr{D}_d(\boldsymbol{x}_{b^m-1})\}$ is a digital (t, m, s)-net.

The search for (t, m, s)-nets has now been reduced to finding suitable matrices $C_1, \ldots, C_s$. Explicit constructions of such matrices are available, see [7, 13].

2.3 Walsh functions

To analyze the RMSE we use the Walsh series expansions of the integrands. In this subsection we recall some basic properties of Walsh functions used in this paper. First we give the definition for the one-dimensional case.

Definition 2 *Let $b \geq 2$ be an integer and represent $k \in \mathbb{N}_0$ in base b, $k = \kappa_{a-1} b^{a-1} + \cdots + \kappa_0$, with $\kappa_i \in \{0, \ldots, b-1\}$. Further let $\omega_b = e^{2\pi i/b}$. Then the kth Walsh function ${}_b\mathrm{wal}_k : [0, 1) \to \left\{1, \omega_b, \ldots, \omega_b^{b-1}\right\}$ in base b is given by*

$${}_b\mathrm{wal}_k(x) = \omega_b^{x_1 \kappa_0 + \cdots + x_a \kappa_{a-1}},$$

for $x \in [0,1)$ with base b representation $x = x_1 b^{-1} + x_2 b^{-2} + \cdots$ (unique in the sense that infinitely many of the x_i are different from $b-1$).

We now extend this definition to the multi-dimensional case.

Definition 3 *For dimension $s \geq 2$, $\boldsymbol{x} = (x_1, \ldots, x_s) \in [0,1)^s$, and $\boldsymbol{k} = (k_1, \ldots, k_s) \in \mathbb{N}_0^s$, we define ${}_b\mathrm{wal}_{\boldsymbol{k}} : [0,1)^s \to \{1, \omega_b, \ldots, \omega_b^{b-1}\}$ by*

$$ {}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{s} {}_b\mathrm{wal}_{k_j}(x_j). $$

As can be seen from the definition, Walsh functions are piecewise constant. For $b = 2$ they are also related to Haar functions.

We need some notation to introduce some further properties of Walsh functions. By $\oplus$ we denote the digitwise addition modulo b, i.e., for $x, y \in [0,1)$ with base b expansions $x = \sum_{i=1}^{\infty} x_i b^{-i}$ and $y = \sum_{i=1}^{\infty} y_i b^{-i}$, we define

$$ x \oplus y = \sum_{i=1}^{\infty} z_i b^{-i}, $$

where $z_i \in \{0, \ldots, b-1\}$ is given by $z_i \equiv x_i + y_i \pmod{b}$, and let $\ominus$ denote the digitwise subtraction modulo b. In the same manner we also define a digitwise addition and digitwise subtraction for nonnegative integers based on the b-adic expansion. For vectors in $[0,1)^s$ or $\mathbb{N}_0^s$, the operators $\oplus$ and $\ominus$ are carried out componentwise. Throughout this paper, we always use base b for the operations $\oplus$ and $\ominus$. Further we call $x \in [0,1)$ a b-adic rational if it can be written in a finite base b expansion. In the following proposition, we summarize some basic properties of Walsh functions.

Proposition 1

1. *For all $k, l \in \mathbb{N}_0$ and all $x, y \in [0,1)$, with the restriction that if x, y are not q-adic rationals, then $x \oplus y$ is not allowed to be a b-adic rational, we have*

$$ {}_b\mathrm{wal}_k(x) \cdot {}_b\mathrm{wal}_l(x) = \mathrm{wal}_{k \oplus l}(x), \quad {}_b\mathrm{wal}_k(x) \cdot {}_b\mathrm{wal}_k(y) = {}_b\mathrm{wal}_k(x \oplus y). $$

2. *We have*

$$ \int_0^1 {}_b\mathrm{wal}_0(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_0^1 {}_b\mathrm{wal}_k(x)\,\mathrm{d}x = 0 \quad \text{if } k > 0. $$

3. *For all $\boldsymbol{k}, \boldsymbol{l} \in \mathbb{N}_0^s$ we have the following orthogonality properties:*

$$ \int_{[0,1)^s} {}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) \overline{{}_b\mathrm{wal}_{\boldsymbol{l}}(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x} = \begin{cases} 1 & \text{if } \boldsymbol{k} = \boldsymbol{l} \\ 0 & \text{otherwise.} \end{cases} $$

4. *For any* $f \in \mathcal{L}_2([0,1)^s)$ *and any* $\boldsymbol{\sigma} \in [0,1)^s$ *we have*

$$\int_{[0,1)^s} f(\boldsymbol{x} \oplus \boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{x} = \int_{[0,1)^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

5. *For* $s \in \mathbb{N}$*, the system* $\{ {}_b\mathrm{wal}_{\boldsymbol{k}} : \boldsymbol{k} = (k_1, \ldots, k_s), k_1, \ldots, k_s \geq 0\}$ *is a complete orthonormal system in* $\mathcal{L}_2([0,1]^s)$*.*

The proofs of 1. – 3. are straightforward, and for a proof of the remaining items see [3] or [7, 21] for more information.

Let $d \geq 1$ and $k_1, \ldots, k_d \in \mathbb{N}_0$. Let $k_i = \kappa_{i,0} + \kappa_{i,1} b + \cdots$, where $\kappa_{i,a} \in \{0, \ldots, b-1\}$ and $\kappa_{i,a} = 0$ for a large enough. To analyze the RMSE, it is convenient to extend the digit interlacing function $\mathscr{D}_d$ to

$$\begin{aligned} \mathscr{D}_d : \mathbb{N}^d &\rightarrow \mathbb{N} \\ (k_1, \ldots, k_d) &\mapsto \sum_{a=0}^{\infty} \sum_{r=1}^{d} \kappa_{r,a} b^{r-1+ad}. \end{aligned}$$

Then we have

$${}_b\mathrm{wal}_{\mathscr{D}_d(k_1,\ldots,k_d)}(\mathscr{D}_d(x_1, \ldots, x_d)) = \prod_{i=1}^{d} {}_b\mathrm{wal}_{k_i}(x_i).$$

2.4 Scrambling

The scrambling algorithm which yields the optimal rate of convergence of the RMSE uses the digit interlacing function and the scrambling introduced by Owen [15, 16, 17], which we describe in the following.

2.4.1 Owen's scrambling

Owen's scrambling algorithm is easiest described for some generic point $\boldsymbol{x} \in [0,1)^s$, with $\boldsymbol{x} = (x_1, \ldots, x_s)$ and $x_i = \xi_{i,1} b^{-1} + \xi_{i,2} b^{-2} + \cdots$. The scrambled point shall be denoted by $\boldsymbol{y} \in [0,1)^s$, where $\boldsymbol{y} = (y_1, \ldots, y_s)$ and $y_i = \eta_{i,1} b^{-1} + \eta_{i,2} b^{-2} + \cdots$. The point $\boldsymbol{y}$ is obtained by applying permutations to each digit of each coordinate of $\boldsymbol{x}$. The permutation applied to $\xi_{i,l}$ depends on $\xi_{i,k}$ for $1 \leq k < l$. Specifically, $\eta_{i,1} = \pi_i(\xi_{i,1})$, $\eta_{i,2} = \pi_{i,\xi_{i,1}}(\xi_{i,2})$, $\eta_{i,3} = \pi_{i,\xi_{i,1},\xi_{i,2}}(\xi_{i,3})$, and in general

$$\eta_{i,k} = \pi_{i,\xi_{i,1},\ldots,\xi_{i,k-1}}(\xi_{i,k}), \tag{1}$$

where $\pi_{i,\xi_{i,1},\ldots,\xi_{i,k-1}}$ is a random permutation of $\{0,\ldots,b-1\}$. We assume that permutations with different indices are chosen mutually independent from each other and that each permutation is chosen with the same probability.

To describe Owen's scrambling, for $1 \le i \le s$ let

$$\Pi_i = \{\pi_{i,\xi_{i,1},\ldots\xi_{i,k-1}} : k \in \mathbb{N}, \xi_{i,1},\ldots,\xi_{i,k-1} \in \{0,\ldots,b-1\}\},$$

where for $k = 1$ we set $\pi_{i,\xi_{i,1},\ldots,\xi_{i,k-1}} = \pi_i$, be a given set of permutations and let $\mathbf{\Pi} = (\Pi_1,\ldots,\Pi_s)$. Then, when applying Owen's scrambling using these permutations to some point $\boldsymbol{x} \in [0,1)^s$, we write $\boldsymbol{y} = \boldsymbol{x}_{\mathbf{\Pi}}$, where $\boldsymbol{y}$ is the point obtained by applying Owen's scrambling to $\boldsymbol{x}$ using the set of permutations $\mathbf{\Pi} = (\Pi_1,\ldots,\Pi_s)$. For $x \in [0,1)$ we drop the subscript i and just write $y = x_\Pi$.

To analyze the RMSE it is also convenient to generalize Owen's scrambling to higher order.

2.4.2 Owen's scrambling of order d

We now describe what we mean by Owen's scrambling of order $d \ge 1$ for a generic point $\boldsymbol{x} \in [0,1)^s$. The scrambled point $\boldsymbol{y} \in [0,1)^s$ is given by

$$\boldsymbol{y} = \mathscr{D}_d\left(\mathscr{D}_d^{-1}(\boldsymbol{x})_{\mathbf{\Pi}}\right),$$

that is, one applies the inverse mapping $\mathscr{D}_d^{-1}$ (see Appendix A for more information on $\mathscr{D}_d$) to the point $\boldsymbol{x}$ to obtain a point $\boldsymbol{z} \in [0,1)^{ds}$, applies Owen's scrambling of Section 2.4.1 to $\boldsymbol{z}$ to obtain a point $\boldsymbol{w} = \boldsymbol{z}_{\mathbf{\Pi}} \in [0,1)^{ds}$ and then use the transformation $\mathscr{D}_d$ to obtain the point $\boldsymbol{y} = \mathscr{D}_d(\boldsymbol{w}) \in [0,1)^s$.

Assuming that the permutations are all chosen with equal probability, then the point $\boldsymbol{y}$ is uniformly distributed in $[0,1)^s$.

Proposition 2 *Let $\boldsymbol{x} \in [0,1)^s$ and let $\mathbf{\Pi}$ be a uniformly and i.i.d. set of permutations. Then $\mathscr{D}_d(\mathscr{D}_d^{-1}(\boldsymbol{x})_{\mathbf{\Pi}})$ is uniformly distributed in $[0,1)^s$, that is, for any Lebesgue measurable set $G \subseteq [0,1)^s$, the probability that $\mathscr{D}_d(\mathscr{D}_d^{-1}(\boldsymbol{x})_{\mathbf{\Pi}})$, denoted by $\mathrm{Prob}[\mathscr{D}_d(\mathscr{D}_d^{-1}(\boldsymbol{x})_{\mathbf{\Pi}})] = \lambda_s(G)$, where λ_s denotes the s-dimensional Lebesgue measure.*

This result follows along the same lines as the proof of [15, Proposition 2].

2.4.3 Owen's lemma of order d

A key result on scrambled nets is Owen's lemma (see [16]) which we now generalize to include the case of scrambling of order d. Let $k \in \mathbb{N}$ have base b representation $k = \kappa_0 + \kappa_1 b + \cdots + \kappa_a b^a$. For $0 \le r < d$ let

$$k_r = \kappa_r b^r + \kappa_{r+d} b^{r+d} + \cdots + \kappa_{a_r} b^{a_r},$$

where $a_r \leq a$ is the largest integer such that d divides $a_r - r$. If $a < r$ we set $k_r = 0$.

For $x = \xi_1 b^{-1} + \xi_2 b^{-2} + \cdots$ and $x' = \xi_1' b^{-1} + \xi_2' b^{-2} + \cdots$ and for $0 \leq r < d$ let β_r be the largest integer such that $\xi_r = \xi_r', \xi_{r+d} = \xi_{r+d}', \ldots, \xi_{r+\beta_r d} = \xi_{r+\beta_r d}'$, and $\xi_{r+(\beta_r+1)d} \neq \xi_{r+(\beta_r+1)d}'$.

Lemma 4 *Let $y, y' \in [0,1)$ be two points obtained by applying Owen's scrambling algorithm of order $d \geq 1$ to the points $x, x' \in [0,1)$.*

(i) If $k \neq k'$, then

$$\mathbb{E}\left[{}_b\mathrm{wal}_k(y) \overline{{}_b\mathrm{wal}_{k'}(y')} \right] = 0.$$

(ii) If $k = k'$ and there exists an $0 \leq r < d$ such that $k_r \geq b^{\beta_r+1}$, then

$$\mathbb{E}\left[{}_b\mathrm{wal}_k(y \ominus y') \right] = 0.$$

(iii) If $k = k'$ and $k_r < b^{\beta_r+1}$ for $0 \leq r < d$, then

$$\mathbb{E}\left[{}_b\mathrm{wal}_k(y \ominus y') \right] = (1-b)^{-v},$$

where

$$v = \left| \{0 \leq r < d : b^{\beta_r} \leq k_r < b^{\beta_r+1}\} \right|.$$

The proof of this result follows immediately from [7, Lemma 13.23].

In the next section we analyze the variance of the estimator $\widehat{I}(f) = \frac{1}{b^m} \sum_{n=0}^{b^m-1} f(\boldsymbol{y}_n)$.

3 Variance of the estimator

Let $f \in L_2([0,1]^s)$ have the following Walsh series expansion

$$f(\boldsymbol{x}) \sim \sum_{\boldsymbol{k} \in \mathbb{N}_0^s} \widehat{f}(\boldsymbol{k}) \, {}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) =: S(\boldsymbol{x}, f). \tag{2}$$

Although we do not necessarily have equality in (2), the completeness of the Walsh function system $\{ {}_b\mathrm{wal}_{\boldsymbol{k}} : \boldsymbol{k} \in \mathbb{N}_0^s\}$ (see [7]) implies that we do have

$$\mathrm{Var}[f] = \sum_{\boldsymbol{k} \in \mathbb{N}_0^s} |\widehat{f}(\boldsymbol{k})|^2 = \mathrm{Var}[S(\cdot, f)]. \tag{3}$$

We estimate the integral $\int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ by

$$\widehat{I}(f) = \frac{1}{b^m} \sum_{n=0}^{b^m-1} f(\boldsymbol{y}_n),$$

where $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{b^m-1} \in [0,1)^s$ is obtained by applying a random Owen scrambling of order d to the digital (t,m,s)-net $P_{b^m} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$ (below we shall assume that there is a digital (t,m,ds)-net $\{\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{b^m-1}\}$ such that $\boldsymbol{x}_n = \mathscr{D}_d(\boldsymbol{z}_n)$ for $0 \le n < b^m$, but for now the assumption that P_{b^m} is a digital (t,m,s)-net is sufficient). From Proposition 2 it follows that

$$\mathbb{E}[\widehat{I}(f)] = \int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

Hence in the following we consider the variance of the estimator $\widehat{I}(f)$ denoted by

$$\operatorname{Var}[\widehat{I}(f)] = \mathbb{E}[(\widehat{I}(f) - \mathbb{E}[\widehat{I}(f)])^2].$$

The following notation is needed for the lemma below. Let $d \ge 1$ and $\boldsymbol{l} = (\boldsymbol{l}_1, \ldots, \boldsymbol{l}_s) \in \mathbb{N}_0^{ds}$, where $\boldsymbol{l}_i = (l_{(i-1)d+1}, \ldots, l_{id})$. Let

$$B_{d,\boldsymbol{l},s} = \{(k_1, \ldots, k_{ds}) \in \mathbb{N}_0^{ds} : \lfloor b^{l_i - 1} \rfloor \le k_i < b^{l_i} \text{ for } 1 \le i \le ds\}.$$

We set

$$\sigma^2_{d,\boldsymbol{l},s}(f) = \sum_{\boldsymbol{k} \in B_{d,\boldsymbol{l},s}} \left| \widehat{f}(\mathscr{D}_d(\boldsymbol{k})) \right|^2.$$

Consider $s = 1$ for a moment. Let $\boldsymbol{l} \in \mathbb{N}_0^d$. Then Lemma 4 implies that for $(k_1, \ldots, k_d), (k_1', \ldots, k_d') \in B_{d,\boldsymbol{l},1}$ we have

$$\begin{aligned} & \mathbb{E}\left[{}_b\mathrm{wal}_{(k_1,\ldots,k_d)}(\mathscr{D}_d^{-1}(x)_{\boldsymbol{\Pi}}) \overline{{}_b\mathrm{wal}_{(k_1,\ldots,k_d)}(\mathscr{D}_d^{-1}(x')_{\boldsymbol{\Pi}})} \right] \\ &= \mathbb{E}\left[{}_b\mathrm{wal}_{(k_1',\ldots,k_d')}(\mathscr{D}_d^{-1}(x)_{\boldsymbol{\Pi}}) \overline{{}_b\mathrm{wal}_{(k_1',\ldots,k_d')}(\mathscr{D}_d^{-1}(x')_{\boldsymbol{\Pi}})} \right]. \end{aligned} \tag{4}$$

Hence, for $s \ge 1$ and $\boldsymbol{l} \in \mathbb{N}_0^{ds}$, choose an arbitrary $\boldsymbol{k} \in B_{d,\boldsymbol{l},s}$, and set

$$\Gamma_{d,\boldsymbol{l}}(P_{b^m}) = \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^{s} \mathbb{E}\left[{}_b\mathrm{wal}_{(k_{d(i-1)+1},\ldots,k_{di})}(\mathscr{D}_d^{-1}(x_{n,i})_{\boldsymbol{\Pi}}) \overline{{}_b\mathrm{wal}_{(k_{d(i-1)+1},\ldots,k_{di})}(\mathscr{D}_d^{-1}(x_{n',i})_{\boldsymbol{\Pi}})} \right].$$

Equation (4) implies that this definition is independent of the particular choice of $\boldsymbol{k} \in B_{d,\boldsymbol{l},s}$. We call $\Gamma_{d,\boldsymbol{l}}(P_{b^m})$ the *gain coefficient (of P_{b^m})(of order d).*

Lemma 5 *Let $d \ge 1$. Let $f \in L_2([0,1]^s)$ and*

$$\widehat{I}(f) = \frac{1}{b^m} \sum_{n=0}^{b^m-1} f(\boldsymbol{y}_n),$$

where $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{b^m-1} \in [0,1)^s$ is obtained by applying a random Owen scrambling of order d to the digital net $P_{b^m} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$. Then

$$\operatorname{Var}[\widehat{I}(f)] = \sum_{\boldsymbol{l} \in \mathbb{N}_0^{ds} \setminus \{\boldsymbol{0}\}} \sigma^2_{d,\boldsymbol{l},s}(f)\, \Gamma_{d,\boldsymbol{l}}(P_{b^m}).$$

Proof. Using the linearity of expectation and Lemma 4 we get

$$\begin{aligned}
\operatorname{Var}[\widehat{I}(f)] &= \mathbb{E}\left[\sum_{\boldsymbol{k},\boldsymbol{k}' \in \mathbb{N}_0^s \setminus \{\boldsymbol{0}\}} \widehat{f}(\boldsymbol{k}) \overline{\widehat{f}(\boldsymbol{k}')} \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} {}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{y}_n) \overline{{}_b\mathrm{wal}_{\boldsymbol{k}'}(\boldsymbol{y}_{n'})}\right] \\
&= \sum_{\boldsymbol{k},\boldsymbol{k}' \in \mathbb{N}_0^s \setminus \{\boldsymbol{0}\}} \widehat{f}(\boldsymbol{k}) \overline{\widehat{f}(\boldsymbol{k}')} \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^s \mathbb{E}\left[{}_b\mathrm{wal}_{k_i}(y_{n,i}) \overline{{}_b\mathrm{wal}_{k_i'}(y_{n',i})}\right] \\
&= \sum_{\boldsymbol{k} \in \mathbb{N}_0^s \setminus \{\boldsymbol{0}\}} |\widehat{f}(\boldsymbol{k})|^2 \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^s \mathbb{E}\left[{}_b\mathrm{wal}_{k_i}(y_{n,i}) \overline{{}_b\mathrm{wal}_{k_i}(y_{n',i})}\right] \\
&= \sum_{\boldsymbol{l} \in \mathbb{N}_0^{ds} \setminus \{\boldsymbol{0}\}} \sum_{\boldsymbol{k} \in B_{d,\boldsymbol{l},s}} |\widehat{f}(\mathscr{D}_d(\boldsymbol{k}))|^2 \\
&\quad \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^s \mathbb{E}\left[{}_b\mathrm{wal}_{(k_{d(i-1)+1},\ldots,k_{di})}(\mathscr{D}_d^{-1}(x_{n,i})_{\boldsymbol{\Pi}} \ominus \mathscr{D}_d^{-1}(x_{n',i})_{\boldsymbol{\Pi}})\right] \\
&= \sum_{\boldsymbol{l} \in \mathbb{N}_0^{ds} \setminus \{\boldsymbol{0}\}} \sigma^2_{d,\boldsymbol{l},s}(f) \Gamma_{d,\boldsymbol{l}}(P_{b^m}).
\end{aligned}$$

Hence the result follows. □

To obtain a bound on the variance $\operatorname{Var}[\widehat{I}(f)]$ we prove bounds on $\sigma_{d,\boldsymbol{l},s}(f)$ and $\Gamma_{d,\boldsymbol{l}}(P_{b^m})$, which we consider in the following two subsections.

3.1 A bound on the gain coefficients of order d

In this section we prove a bound on $\Gamma_{d,\boldsymbol{l}}(P_{b^m})$, where the point set is a digital (t,m,s)-net as constructed in [5].

Lemma 6 *Let $\{\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{b^m-1}\}$ be a digital (t,m,ds)-net over $\mathbb{Z}_b$. Let $\boldsymbol{x}_n = \mathscr{D}_d(\boldsymbol{z}_n)$ for $0 \le n < b^m$. Then the gain coefficients of order d for the digital net $P_{b^m} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$ satisfy*

$$\Gamma_{d,\boldsymbol{l}}(P_{b^m}) \le \begin{cases} 0 & \text{if } |\boldsymbol{l}|_1 \le m-t, \\ b^{|q|-|\boldsymbol{l}|_1} & \text{if } m-t < |\boldsymbol{l}|_1 \le m-t+|q|, \\ b^{-m+t} & \text{if } |\boldsymbol{l}|_1 > m-t+|q|. \end{cases}$$

Proof. Let $\boldsymbol{k} = (k_1, \ldots, k_{ds})$ and $\boldsymbol{l} = (\boldsymbol{l}_q, \boldsymbol{0})$ for some $q \subseteq \{1, \ldots, s\}$. Then from the proof of [7, Corollary 13.7] and [7, Lemma 13.8] it follows that

$$\begin{aligned}
\Gamma_{d,\boldsymbol{l}}(P_{b^m}) &= \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^{s} \mathbb{E}\left[{}_b\mathrm{wal}_{(k_{d(i-1)+1},\ldots,k_{di})}(\mathscr{D}_d^{-1}(x_{n,i})_{\boldsymbol{\Pi}}) \overline{{}_b\mathrm{wal}_{(k'_{d(i-1)+1},\ldots,k'_{di})}(\mathscr{D}_d^{-1}(x_{n',i})_{\boldsymbol{\Pi}})} \right] \\
&= \frac{1}{b^{2m}} \sum_{n,n'=0}^{b^m-1} \prod_{i=1}^{s} \mathbb{E}\left[{}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{z}_n)_{\boldsymbol{\Pi}}) \overline{{}_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{z}_{n'})_{\boldsymbol{\Pi}})} \right] \\
&= \begin{cases} 0 & \text{if } |\boldsymbol{l}|_1 \le m - t, \\ b^{|q| - |\boldsymbol{l}|_1} & \text{if } m - t < |\boldsymbol{l}|_1 \le m - t + |q|, \\ b^{-m+t} & \text{if } |\boldsymbol{l}|_1 > m - t + |q|. \end{cases}
\end{aligned}$$

Hence the result follows. $\square$

3.2 Higher order variation

In this subsection we state a bound on $\sigma_{d,\boldsymbol{l},s}(f)$. The rate of decay of $\sigma_{d,\boldsymbol{l},s}(f)$ depends on the smoothness of the function f. We measure the smoothness using a variation based on finite differences, which we introduce in the following. Since the smoothness of the function f may be unknown, we cannot assume that we can choose d to be the smoothness. Hence, in the following we use α to denote the smoothness of the integrand f.

3.2.1 Finite differences

We use a slight variation from classical finite differences. Let $f : [0,1] \to \mathbb{R}$ and let $z_1, z_2, \ldots \in (-1,1)$ be a sequence of numbers. Then we define $\Delta_0(x)f = f(x)$ and for $\alpha \ge 1$ we set

$$\Delta_\alpha(x; z_1, \ldots, z_\alpha)f = \Delta_{\alpha-1}(x + z_\alpha; z_1, \ldots, z_{\alpha-1})f - \Delta_{\alpha-1}(x; z_1, \ldots, z_{\alpha-1})f.$$

For instance, we have

$$\begin{aligned}
\Delta_1(x; z_1)f &= f(x + z_1) - f(x), \\
\Delta_2(x; z_1, z_2)f &= f(x + z_1 + z_2) - f(x + z_2) - f(x + z_1) + f(x),
\end{aligned}$$

and in general

$$\Delta_\alpha(x; z_1, \ldots, z_\alpha)f = \sum_{v \subseteq \{1,\ldots,\alpha\}} (-1)^{|v|} f\left(x + \sum_{i \in v} z_i\right),$$

where $|v|$ denotes the number of elements in v. We always assume that $x + \sum_{i\in v} z_i \in [0,1]$ for all $v \subseteq \{1,\dots,\alpha\}$.

If f is α times continuously differentiable, then the mean value theorem implies that

$$\Delta_\alpha(x; z_1,\dots,z_\alpha) f = z_\alpha \Delta_{\alpha-1}(\zeta_1; z_1,\dots,z_{\alpha-1}) \frac{\mathrm{d}f}{\mathrm{d}x},$$

where $\min(x, x+z_\alpha) \le \zeta_1 \le \max(x, x+z_\alpha)$. By induction, it then follows that

$$\Delta_\alpha(x; z_1,\dots,z_\alpha) f = z_1 \cdots z_\alpha \frac{\mathrm{d}^\alpha f}{\mathrm{d}x^\alpha}(\zeta_\alpha),$$

where

$$x + \min_{v\subseteq\{1,\dots,\alpha\}} \sum_{i\in v} z_i \le \zeta_\alpha \le x + \max_{v\subseteq\{1,\dots,\alpha\}} \sum_{i\in v} z_i.$$

We generalize the difference operator to functions $f : [0,1]^s \to \mathbb{R}$. Let $\alpha > 0$ be a nonnegative integer. Let $\Delta_{i,\alpha}$ be the one-dimensional difference operator Δ_α applied to the ith coordinate of f. For $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_s) \in \{0,\dots,\alpha\}^s$ and $1 \le i \le s$ let $z_{i,1},\dots,z_{i,\alpha_i} \in (-1,1)$. Then we define

$$\begin{aligned}
&\Delta_{\boldsymbol{\alpha}}(\boldsymbol{x}; (z_{1,1},\dots,z_{1,\alpha_1}),\dots,(z_{s,1},\dots,z_{s,\alpha_s})) f \\
&= \Delta_{1,\alpha_1}(x_1; z_{1,1},\dots,z_{1,\alpha_1}) \cdots \Delta_{s,\alpha_s}(x_s; z_{s,1},\dots,z_{s,\alpha_s}) f \\
&= \sum_{v_1\subseteq\{1,\dots,\alpha_1\}} \cdots \sum_{v_s\subseteq\{1,\dots,\alpha_s\}} (-1)^{|v_1|+\cdots+|v_s|} f\left(x_1 + \sum_{i_1\in v_1} z_{1,i_1},\dots,x_s + \sum_{i_s\in v_s} z_{s,i_s}\right).
\end{aligned}$$

If f has continuous mixed partial derivatives up to order α in each variable, then, as for the one-dimensional case, we have

$$\Delta_{\boldsymbol{\alpha}}(\boldsymbol{x}, (z_{1,1},\dots,z_{1,\alpha_1}),\dots,(z_{s,1},\dots,z_{s,\alpha_s})) f = \prod_{i=1}^{s} \prod_{r_i=1}^{\alpha_i} z_{i,r_i} \; \frac{\partial^{\alpha_1+\cdots+\alpha_s} f}{\partial x_1^{\alpha_1} \cdots \partial x_s^{\alpha_s}}(\zeta_{1,\alpha_1},\dots,\zeta_{s,\alpha_s}), \quad (5)$$

where we set $\prod_{r_i=1}^{\alpha_i} z_{i,r_i} = 1$ for $\alpha_i = 0$ and where

$$x_i + \min_{v\subseteq\{1,\dots,\alpha_i\}} \sum_{r\in v} z_{i,r} \le \zeta_{i,\alpha_i} \le x_i + \max_{v\subseteq\{1,\dots,\alpha_i\}} \sum_{r\in v} z_{i,r}$$

for $1 \le i \le s$. Again we assume that $x_i + \sum_{r\in v} z_{i,r} \in [0,1]$ for all $v \subseteq \{1,\dots,\alpha_i\}$, $\zeta_{i,\alpha_i} \in [0,1]$ for all $0 \le \alpha_i \le \alpha$ and $1 \le i \le s$.

3.2.2 Variation

Let $f : [0,1]^s \to \mathbb{R}$ and $\alpha > 0$ be a nonnegative integer. Let $J = \prod_{i=1}^{\alpha s}[\frac{a_i}{b^{l_i}}, \frac{a_i+1}{b^{l_i}})$, with $0 \le a_i < b^{l_i}$ and $l_i \in \mathbb{N}$ for $1 \le i \le \alpha s$. Apart from at most a countable number of points, the set $\mathscr{D}_\alpha(J)$ is the product of a union of intervals. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s) \in \{1, \ldots, \alpha\}^s$. Then we define the generalized Vitali variation by

$$V_{\boldsymbol{\alpha}}^{(s)}(f) = \sup_{\mathcal{P}} \left(\sum_{J \in \mathcal{P}} \mathrm{Vol}(\mathscr{D}_\alpha(J)) \sup \left| \frac{\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_s) f}{\prod_{i=1}^s \prod_{r=1}^{\alpha_i} z_{i,r}} \right|^2 \right)^{1/2}, \tag{6}$$

where the first supremum $\sup_{\mathcal{P}}$ is extended over all partitions of $[0,1)^{\alpha s}$ into subcubes of the form $J = \prod_{i=1}^{\alpha s}[\frac{a_i}{b^{l_i}}, \frac{a_i+1}{b^{l_i}})$ with $0 \le a_i < b^{l_i}$ and $l_i \in \mathbb{N}$ for $1 \le i \le \alpha s$, and the second supremum is taken over all $\boldsymbol{t} \in \mathscr{D}_\alpha(J)$ and $\boldsymbol{z}_i = (z_{i,1}, \ldots, z_{i,\alpha_i})$ with $z_{i,r} = \tau_{i,r} b^{-\alpha(l_i-1)-r}$ where $\tau_{i,r} \in \{1-b, \ldots, b-1\} \setminus \{0\}$ for $1 \le r \le \alpha_i$ and $1 \le i \le s$ and such that all the points at which f is evaluated in $\Delta_\alpha(\boldsymbol{t}; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_s)$ are in $\mathscr{D}_\alpha(\prod_{i=1}^{\alpha s}[b^{-l_i+1}\lfloor a_i/b \rfloor, b^{-l_i+1}(\lfloor a_i/b \rfloor + 1))$.

In Appendix A it is shown that $\mathrm{Vol}(\mathscr{D}_\alpha(J)) = \mathrm{Vol}(J)$, the volume (i.e. Lebesgue measure) of J. Hence, if the partial derivative $\frac{\partial^{\alpha_1+\cdots+\alpha_s} f}{\partial x_1^{\alpha_1} \cdots \partial x_s^{\alpha_s}}$ are continuous for a given $(\alpha_1, \ldots, \alpha_s) \in \{1, \ldots, \alpha\}^s$, then it can be shown that (5) and the mean value theorem imply that the sum (6) is a Riemann sum for the integral

$$V_{\boldsymbol{\alpha}}^{(s)}(f) = \left(\int_{[0,1]^s} \left| \frac{\partial^{\alpha_1+\cdots+\alpha_s} f}{\partial x_1^{\alpha_1} \cdots \partial x_s^{\alpha_s}}(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x} \right)^{1/2}.$$

For $\emptyset \ne u \subseteq \{1, \ldots, s\}$, let $|u|$ denote the number of elements in the set u and let $V_{\boldsymbol{\alpha}_u}^{(|u|)}(f_u; u)$ be the generalized Vitali variation with coefficient $\boldsymbol{\alpha}_u \in \{1, \ldots, \alpha\}^{|u|}$ of the $|u|$-dimensional function

$$f_u(\boldsymbol{x}_u) = \int_{[0,1]^{s-|u|}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}_{\{1,\ldots,s\} \setminus u}.$$

For $u = \emptyset$ we have $f_\emptyset = \int_{[0,1]^s} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ and we define $V_{\boldsymbol{\alpha}}^{(|\emptyset|)}(f_\emptyset; \emptyset) = |f_\emptyset|$.

Then

$$V_\alpha(f) = \left(\sum_{u \subseteq \{1,\ldots,s\}} \sum_{\boldsymbol{\alpha} \in \{1,\ldots,\alpha\}^{|u|}} \left(V_{\boldsymbol{\alpha}}^{(|u|)}(f_u; u) \right)^2 \right)^{1/2}$$

is called the generalized Hardy and Krause variation of f of order α. A function f for which $V_\alpha(f)$ is finite is said to be of bounded variation (of order α).

If the partial derivatives $\frac{\partial^{\alpha_1+\cdots+\alpha_s} f}{\partial x_1^{\alpha_1}\cdots\partial x_s^{\alpha_s}}$ are continuous for all $(\alpha_1,\ldots,\alpha_s) \in \{0,\ldots,\alpha\}^s$, then variation coincides with the norm

$$V_\alpha(f) = \left(\sum_{u\subseteq\{1,\ldots,s\}} \sum_{\boldsymbol{\alpha}\in\{1,\ldots,\alpha\}^{|u|}} \int_{[0,1]^{|u|}} \left| \int_{[0,1]^{s-|u|}} \frac{\partial^{\sum_{i\in u}\alpha_i} f}{\prod_{i\in u}\partial x_i^{\alpha_i}} \,\mathrm{d}\boldsymbol{x}_{\{1,\ldots,s\}\setminus u} \right|^2 \mathrm{d}\boldsymbol{x}_u \right)^{1/2}.$$

3.2.3 The decay of the Walsh coefficients for functions of bounded variation

The following lemma gives a bound on $\sigma_{d,\boldsymbol{l},s}(f)$ for functions f of bounded variation of order α.

Lemma 7 *Let $\alpha, d \in \mathbb{N}$. Let $f : [0,1]^s \to \mathbb{R}$ with $V_\alpha(f) < \infty$. Let $b \ge 2$ be an integer. Let $\boldsymbol{l} = (l_1,\ldots,l_{ds}) \in \mathbb{N}_0^{ds}$ and let $K = \{i \in \{1,\ldots,ds\} : l_i > 0\}$. Let $K_i = K \cap \{(i-1)d+1,\ldots,id\}$ and $\alpha_i = \min(\alpha,|K_i|)$ for $1 \le i \le s$. Let $\gamma_j' = (b-1)b^{-j+(i-1)d-(l_j-1)d}$ for $j \in K_i$ and $1 \le i \le s$. Let $\gamma_{i,1} < \gamma_{i,2} < \cdots < \gamma_{i,\alpha_i}$ for $1 \le i \le s$ be such that $\{\gamma_{i,1},\ldots,\gamma_{i,\alpha_i}\} = \{\gamma_j : j \in K_i\}$, that is, $\{\gamma_{i,j} : 1 \le j \le \alpha_i\}$ is just a reordering of the elements of the set $\{\gamma_j : j \in K_i\}$. Set $\gamma(\boldsymbol{l}) = \prod_{i=1}^s \prod_{j=1}^{\alpha_i} \gamma_{i,j}$. Then*

$$\sigma_{d,\boldsymbol{l},s}(f) \le 2^{s\max(d-\alpha,0)}\gamma(\boldsymbol{l})V_\alpha(f).$$

The proof of this result is technical and is therefore deferred to Appendix B.

3.3 Convergence rate

We can now use Lemmas 5, 6 and 7 to prove the main result of the paper.

Theorem 8 *Let $\alpha, d \in \mathbb{N}$. Let $f : [0,1]^s \to \mathbb{R}$ satisfy $V_\alpha(f) < \infty$. Let*

$$\widehat{I}(f) = \frac{1}{b^m}\sum_{n=0}^{b^m-1} f(\boldsymbol{y}_n),$$

where $\boldsymbol{y}_0,\ldots,\boldsymbol{y}_{b^m-1} \in [0,1)^s$ with $\boldsymbol{y}_n = \mathscr{D}_d((\boldsymbol{x}_n)_{\boldsymbol{\Pi}})$ and $\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{b^m-1} \in [0,1)^{ds}$ is a digital (t,m,ds)-net and the permutations in $\boldsymbol{\Pi}$ are chosen uniformly and i.i.d.. Then

$$\mathrm{Var}[\widehat{I}(f)] \le C_{b,s,\alpha}V_\alpha(f)\frac{(m-t)^{\min(\alpha,d)s+s}}{b^{-(2\min(\alpha,d)+1)(m-t)}},$$

where $C_{b,s,\alpha} > 0$ is a constant which depends only on α, b, d, s, but not on m.

Proof. Let $d \leq \alpha$. Then from Lemmas 5, 6, 7 and the fact that $V_d(f) \leq V_\alpha(f)$ we obtain that

$$\begin{aligned}
\mathrm{Var}[\widehat{I}(f)] &\leq V_\alpha(f)(b-1)^{2ds} b^{s+d(d-1)} b^{-(m-t+1)} \sum_{\boldsymbol{l} \in \mathbb{N}_0^{ds}, |\boldsymbol{l}|_1 > m-t} b^{-2d|\boldsymbol{l}|_1} \\
&\leq V_\alpha(f)(b-1)^{2ds} b^{s+d(d-1)} b^{-(m-t+1)} \sum_{k=m-t+1}^{\infty} b^{-2dk} \binom{k+ds-1}{ds-1} \\
&\leq V_\alpha(f)(b-1)^{2ds} (b^{2d}-1)^{-ds} b^{2d^2 s+s+d(d-1)} b^{-(2d+1)(m-t+1)} \binom{m-t+ds}{ds-1}
\end{aligned}$$

where we used [7, Lemma 13.24]. Since

$$\binom{m-t+ds}{ds-1} = \frac{(m-t+ds)\cdots(m-t+2)}{(ds-1)\cdot\cdots\cdot 1} \leq (m-t+2)^{ds-1}$$

we obtain

$$\mathrm{Var}[\widehat{I}(f)] \leq C_{\alpha,b,d,s} V_\alpha(f) b^{-(2d+1)(m-t)} (m-t+2)^{ds-1},$$

for some constant $C_{\alpha,,d,s} > 0$ which depends only on α, b, d, s.

Let now $d > \alpha$. In the following we sum over all $\boldsymbol{l} = (\boldsymbol{l}_1, \ldots, \boldsymbol{l}_s) \in \mathbb{N}_0^{ds}$, where $\boldsymbol{l}_i = (l_{(i-1)d+1}, \ldots, l_{id})$, and such that $l_1 + \cdots + l_{ds} > m-t$. Let $l'_{(i-1)d+1} \geq l'_{(i-1)d+2} \geq \cdots \geq l'_{id}$ be such that $\{l'_{(i-1)d+1}, \ldots, l'_{id}\} = \{l_{(i-1)d+1}, \ldots, l_{id}\}$, that is, the l'_i are just a reordering of the elements l_i. There are at most $(d!)^s$ reorderings which yield the same $\boldsymbol{l}'_1, \ldots, \boldsymbol{l}'_s$. Then we have

$$\prod_{j=1}^{\alpha_i} \gamma_{i,j} \leq (b-1)^{\alpha_i} b^{(d-1)+(d-2)+\cdots+(d-\alpha_i)} \prod_{j=1}^{\alpha_i} b^{-dl'_i} \leq (b-1)^{\alpha} b^{d(d-1)/2} b^{-d\sum_{j=1}^{\alpha_i} l'_{(i-1)d+j}}.$$

Hence we have

$$\mathrm{Var}[\widehat{I}(f)] \leq V_\alpha(f) 4^{s(d-\alpha)} (b-1)^{2\alpha} b^{s+d(d-1)} (d!)^s b^{-(m-t+1)} \sum_{\substack{\boldsymbol{l} \in \mathbb{N}_0^{ds}, |\boldsymbol{l}|_1 > m-t \\ \boldsymbol{l} \text{ ordered}}} b^{-2d\sum_{i=1}^{s}\sum_{j=1}^{\alpha} l_{(i-1)d+j}}, \quad (7)$$

where $\boldsymbol{l} = (l_1, \ldots, l_{ds})$ ordered means that $l_{(i-1)d+1} \geq \cdots \geq l_{id}$ for $1 \leq i \leq s$. Hence we have

$$m-t < l_1 + \cdots + l_{ds} \leq \frac{d}{\alpha} \sum_{i=1}^{s} \sum_{j=1}^{\alpha} l_{(i-1)d+j}.$$

Let now $k_i = l_{(i-1)d+1} + \cdots + l_{(i-1)d+\alpha}$. Then $k_i \ge \alpha l_{(i-1)d+j}$ for $\alpha < j \le d$ and $k_1 + \cdots + k_s \ge \alpha(m-t)/d$. Hence

$$\begin{aligned}
& \sum_{\substack{\boldsymbol{l} \in \mathbb{N}_0^{ds}, |\boldsymbol{l}|_1 > m-t \\ \boldsymbol{l} \text{ ordered}}} b^{-2d \sum_{i=1}^s \sum_{j=1}^\alpha l_{(i-1)d+j}} \\
\le & \sum_{k_1,\ldots,k_s \in \mathbb{N}_0, k_1+\cdots+k_s > \alpha(m-t)/d} b^{-2d(k_1+\cdots+k_s)} \prod_{i=1}^s \binom{k_i + \alpha - 1}{\alpha - 1} \left(\frac{k_i}{\alpha} + 1\right)^{s(d-\alpha)} \\
\le & \sum_{p_1,\ldots,p_s \in \mathbb{N}_0, p_1+\cdots+p_s > \alpha(m-t)} b^{-2(p_1+\cdots+p_s)} \prod_{i=1}^s \binom{\lceil p_i/d \rceil + \alpha - 1}{\alpha - 1} \left(\left\lceil \frac{p_i}{\alpha d} \right\rceil + 1\right)^{s(d-\alpha)} \\
\le & \sum_{p_1,\ldots,p_s \in \mathbb{N}_0, p_1+\cdots+p_s > \alpha(m-t)} b^{-2(p_1+\cdots+p_s)} \left(\frac{p_i}{d} + 2\right)^{sd} \\
\le & \sum_{p=\alpha(m-t)+1}^{\infty} b^{-2p} \binom{p+s-1}{s-1} \left(\frac{p}{d} + 2\right)^{sd} \\
\le & \sum_{p=\alpha(m-t)+1}^{\infty} b^{-2p} (p+2)^{sd+s-1} \\
\le & \; b^{-2\alpha(m-t)} (\alpha(m-t)+2)^{sd+s} (s(d+1)-1) \times \\
& \max(1, (s(d+1)-1)^{s(d+1)-1} (\alpha(m-t)+1)^{-(s(d+1)-1)} (\log b)^{-(s(d+1)-1)}).
\end{aligned}$$

Thus the result follows from (7). □

4 Discussion

In this paper we have extended the results of [17, 19], by introducing an algorithm and proving that this algorithm can take advantage of the smoothness of the integrand α, where $\alpha \in \mathbb{N}$ can be arbitrarily large. Theorem 8 shows the convergence rate of the standard deviation of the estimator $\widehat{I}(f)$ of $\mathcal{O}(N^{-\min(\alpha,d)-1/2} (\log N)^{s \min(\alpha+1,d+1)/2})$. The numerical results in Section 1.2 using some toy examples also exhibit this rate of convergence. The upper bound is best possible (apart from the power of the $\log N$ factor), since there is also a lower bound on the standard deviation, see [14].

The improvement in the rate of convergence in [19] has been obtained by using variance reduction techniques. Conversely, one might now ask whether the methods developed here can be used to obtain new variance reduction techniques. (Some similarities between this

approach and antithetic sampling can be found in [6].) This is an open question for future research.

Since the classical scrambling by Owen [15] is computationally to expensive, variations of this scrambling scheme have been introduced which can easily be implemented. Matoušek [11, 12] describes an alternative scrambling which uses less permutations and is therefore easier to implement, see also [10]. Another scrambling scheme which can be implemented is by Tezuka and Faure [20]. See also [8, 18, 19] for overviews of various scramblings. The idea is to reduce the number of permutations required such that Owen's lemma still holds. Since the proof of Lemma 4 follows along the same lines as the proof of Owen's lemma, the simplified scramblings mentioned above also apply here.

The only alternative algorithm which achieves the same convergence rate of the RMSE as proven here is based on using an approximation $A(f)$ to the integrand f and then applying MC to $A(f) - f$. The integral is then approximated by $\widehat{I}(A(f) - f) + \int_{[0,1]^s} A(f)(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ where $\int_{[0,1]^s} A(f)(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ can be calculated analytically. See [1, 9] for details.

5 Appendix A: Properties of the digit interlacing function

The digit interlacing function has several properties which we investigate in the following and which we use below.

Lemma 9 *Let $d > 1$. Then the mapping $\mathscr{D}_d : [0,1)^{ds} \to [0,1)^s$ is injective but not surjective.*

Proof. It suffices to show the result for $s = 1$. First note that the digit expansion of $\mathscr{D}_d(x_1, \ldots, x_d)$ is never of the form $c_1 b^{-1} + \cdots + c_j b^{-j+1} + (b-1)b^{-j} + (b-1)b^{-j-d} + (b-1b^{-j-2d} + \cdots$, since this would imply that there is a x_{j_0}, $1 \le j_0 \le d$, which is a b-adic rational. But in this case we use the finite digit expansions of x_{j_0} and hence no vector $(x_1, \ldots, x_d)$ gets mapped to this real number. Thus $\mathscr{D}_d$ is not surjective.

To show that $\mathscr{D}_d$ is injective, let $(x_1, \ldots, x_d) \neq (y_1, \ldots, y_d) \in [0,1)^d$. Hence there exists an $1 \le i \le d$ such that $x_i \neq y_i$, and hence there is a $k \ge 1$ such that $x_{i,k} \neq y_{i,k}$, where $x_i = x_{i,1} b^{-1} + x_{i,2} b^{-2} + \cdots$ and $y_i = y_{i,1} b^{-1} + y_{i,2} b^{-2} + \cdots$ (and where we use the finite expansions for b-adic rationals). Thus the digit expansions of $\mathscr{D}_d(x_1, \ldots, x_d)$ and $\mathscr{D}_d(y_1, \ldots, y_d)$ differ at least at one digit and hence $\mathscr{D}_d(x_1, \ldots, x_d) \neq \mathscr{D}_d(y_1, \ldots, y_d)$. $\square$

(Notice that a countable number of elements could be excluded from the set $[0,1)^s$ such that $\mathscr{D}_d$ becomes bijective.)

Lemma 10 *Let $d \geq 1$ and $J = \prod_{i=1}^{ds}[a_i, b_i) \subseteq [0,1]^{ds}$ with $a_i \leq b_i$ for $1 \leq i \leq ds$. Let λ_n denote the Lebesgue measure on $\mathbb{R}^n$. Then $\lambda_{ds}(J) = \lambda_s(\mathscr{D}_d(J))$.*

Proof. The result is trivial for $d = 1$. Let now $d > 1$.

Consider $s = 1$. Let $J = \prod_{i=1}^{d}[a_i b^{-\nu_i}, (a_i+1)b^{-\nu_i})$, where $0 \leq a_i < b^{\nu_i}$ is an integer and

$$\frac{a_i}{b^{\nu_i}} = \frac{a_{i,1}}{b} + \frac{a_{i,2}}{b^2} + \cdots + \frac{a_{i,\nu_i}}{b^{\nu_i}}$$

for some integers $\nu_i \geq 0$. Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_\alpha)$, $|\boldsymbol{\nu}|_\infty = \max_{1 \leq i \leq s} \nu_i$ and $|\boldsymbol{\nu}|_1 = \nu_1 + \cdots + \nu_s$. Then $\lambda_d(J) = b^{-|\boldsymbol{\nu}|_1}$.

Consider now $\mathscr{D}_d(J)$. Let $0 \leq c < b^{d|\boldsymbol{\nu}|_\infty}$ and

$$c b^{-d|\boldsymbol{\nu}|_\infty} = \frac{c_1}{b} + \frac{c_2}{b^2} + \cdots + \frac{c_{d|\boldsymbol{\nu}|_\infty}}{b^{d|\boldsymbol{\nu}|_\infty}},$$

with $c_1, \ldots, c_{d|\boldsymbol{\nu}|_\infty} \in \{0, \ldots, b-1\}$. We have

$$\mathscr{D}_d(J) = \bigcup \left[\frac{c}{b^{d|\boldsymbol{\nu}|_\infty}}, \frac{c+1}{b^{d|\boldsymbol{\nu}|_\infty}}\right),$$

where the union is over all c with expansion as above and where $c_1, \ldots, c_{d|\boldsymbol{\nu}|_\infty} \in \{0, \ldots, b-1\}$ with the restriction that $a_{i,k} = c_{(k-1)d+i}$ for $1 \leq k \leq \nu_i$ and $1 \leq i \leq d$. Hence there are $d|\boldsymbol{\nu}|_\infty - |\boldsymbol{\nu}|_1$ digits c_j free to choose. Therefore

$$\lambda_1(\mathscr{D}_d(J)) = \lambda_1\left(\left[\frac{c}{b^{d|\boldsymbol{\nu}|_\infty}}, \frac{c+1}{b^{d|\boldsymbol{\nu}|_\infty}}\right)\right) b^{d|\boldsymbol{\nu}|_\infty - |\boldsymbol{\nu}|_1} = b^{-|\boldsymbol{\nu}|_1}.$$

Therefore the result holds for intervals of the form J.

It follows that the result holds for intervals of the form $J = \prod_{i=1}^{ds}[a_i b^{-\nu_i}, (a_i+1)b^{-\nu_i})$, since this interval is simply a product of the previously considered intervals.

Let now $J = \prod_{i=1}^{ds}[a_i, b_i) \subseteq [0,1)^{ds}$, with $a_i < b_i$ for $1 \leq i \leq ds$, be an arbitrary interval. Since this interval can be written as a disjoint union of the elementary intervals used above, the result also holds for these intervals.

Let $\emptyset \neq I \subseteq \{1, \ldots, ds\}$ and $a_i = b_i$ for $i \in I$. Then $\lambda_{ds}(J) = 0$. On the other hand, define

$$b_i' = \begin{cases} a_i + b^{-\nu} & \text{for } i \in I, \\ b_i & \text{otherwise,} \end{cases}$$

where ν is large enough such that $b_i' < 1$ for all $1 \leq i \leq ds$. Set $J' = \prod_{i=1}^{ds}[a_i, b_i')$. Then

$$0 \leq \lambda_s(\mathscr{D}_d(J)) \leq \lambda_s(\mathscr{D}_d(J')) = \lambda_{ds}(J') \leq b^{-\nu} \to 0 \quad \text{as } \nu \to \infty.$$

Hence $\lambda_s(\mathscr{D}_d(J)) = 0$. $\square$

6 Appendix B: Proof of Lemma 7

Assume first that $d \ge \alpha$. Let $\boldsymbol{l} = (l_1, \ldots, l_{ds}) \in \mathbb{N}_0^{ds}$ and let $K = \{i \in \{1, \ldots, ds\} : l_i > 0\}$. Let $K_i = K \cap \{(i-1)d+1, \ldots, (i-1)d+d\}$. First assume that $K_i \neq \emptyset$ for $i = 1, \ldots, s$.

Let $\boldsymbol{l} - \boldsymbol{1}_K = ((l_1 - 1)_+, \ldots, (l_{ds} - 1)_+) \in \mathbb{N}_0^{ds}$, where $(x)_+ = \max(x, 0)$. Let $A_{\boldsymbol{l}} = \{\boldsymbol{a} = (a_1, \ldots, a_{ds}) \in \mathbb{N}_0^{ds} : 0 \le a_i < b^{l_i} \text{ for } 1 \le i \le ds\}$ and

$$[\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}}) := \prod_{i=1}^{ds} [a_i b^{-l_i}, (a_i + 1) b^{-l_i}).$$

Let $\boldsymbol{q} = (q_1, \ldots, q_{\alpha s})$, where $q_i = \lfloor a_i / b \rfloor$. In the following we write $[\boldsymbol{q} b^{-\boldsymbol{l}+\boldsymbol{1}}, (\boldsymbol{q}+\boldsymbol{1}) b^{-\boldsymbol{l}+\boldsymbol{1}})$ for $\prod_{i=1}^{\alpha s} [b^{-l_i+1} \lfloor a_i/b \rfloor, b^{-l_i+1}(\lfloor a_i/b \rfloor + 1)$. Further let

$$\mathscr{D}_d([\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}})) = \{\mathscr{D}_d(\boldsymbol{x}) \in [0,1)^s : \boldsymbol{x} \in [\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}})\}.$$

Let $\boldsymbol{x} \in \mathscr{D}_d([\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}}))$, then

$$\begin{aligned}
\sum_{\boldsymbol{k} \in A_{\boldsymbol{l}}} \widehat{f}(\mathscr{D}_d(\boldsymbol{k}))\, {}_b\mathrm{wal}_{\mathscr{D}_d(\boldsymbol{k})}(\boldsymbol{x}) &= \int_{[0,1]^s} f(\boldsymbol{t}) \sum_{\boldsymbol{k} \in A_{\boldsymbol{l}}} {}_b\mathrm{wal}_{\mathscr{D}_d(\boldsymbol{k})}(\boldsymbol{x} \ominus \boldsymbol{t}) \,\mathrm{d}\boldsymbol{t} \\
&= b^{|\boldsymbol{l}|_1} \int_{\mathscr{D}_d([\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}}])} f(\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}.
\end{aligned}$$

For $\boldsymbol{l} \in \mathbb{N}_0^{ds}$ and $\boldsymbol{a} \in A_{\boldsymbol{l}}$ let

$$c_{\boldsymbol{l},\boldsymbol{a}} = \int_{\mathscr{D}_d([\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}}])} f(\boldsymbol{t}) \,\mathrm{d}\boldsymbol{t}.$$

For $\boldsymbol{x} \in \mathscr{D}_d([\boldsymbol{a} b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1}) b^{-\boldsymbol{l}}))$ let

$$\begin{aligned}
g(\boldsymbol{x}) &:= \sum_{u \subseteq K} (-1)^{|u|} \sum_{\boldsymbol{k} \in A_{\boldsymbol{l} - (\boldsymbol{1}_u, \boldsymbol{0})}} \widehat{f}(\mathscr{D}_d(\boldsymbol{k}))\, {}_b\mathrm{wal}_{\mathscr{D}_d(\boldsymbol{k})}(\boldsymbol{x}) \\
&= \sum_{u \subseteq K} (-1)^{|u|} b^{|\boldsymbol{l} - (\boldsymbol{1}_u, \boldsymbol{0})|_1} c_{\boldsymbol{l} - (\boldsymbol{1}_u, \boldsymbol{0}), (\lfloor \boldsymbol{a}_u / b \rfloor, \boldsymbol{a}_{\{1,\ldots,ds\} \setminus u})},
\end{aligned}$$

where $(\lfloor \boldsymbol{a}_u / b \rfloor, \boldsymbol{a}_{\{1,\ldots,ds\} \setminus u})$ is the vector whose ith coordinate is $\lfloor a_i / b \rfloor$ if $i \in u$ and a_i if $i \in \{1, \ldots, ds\} \setminus u$.

Using Plancherel's identity we obtain

$$\sigma_{d,\boldsymbol{l},s,\boldsymbol{r}}^2(f) = \sum_{u \subseteq K} (-1)^{|u|} \sum_{\boldsymbol{k} \in A_{\boldsymbol{l} - (\boldsymbol{1}_u, \boldsymbol{0})}} |\widehat{f}(\mathscr{D}_d(\boldsymbol{k}))|^2 = \int_0^1 |g(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} b^{-|\boldsymbol{l}|_1} \left| \sum_{u \subseteq K} (-1)^{|u|} b^{|\boldsymbol{l}-(\mathbf{1}_u,\mathbf{0})|_1} c_{\boldsymbol{l}-(\mathbf{1}_u,\mathbf{0}),(\lfloor \boldsymbol{a}_u/b \rfloor, \boldsymbol{a}_{\{1,\ldots,ds\}\setminus u})} \right|^2 \\
&= b^{|\boldsymbol{l}|_1} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} \left| \sum_{u \subseteq K} (-1)^{|u|} b^{-|u|} c_{\boldsymbol{l}-(\mathbf{1}_u,\mathbf{0}),(\lfloor \boldsymbol{a}_u/b \rfloor, \boldsymbol{a}_{\{1,\ldots,ds\}\setminus u})} \right|^2 .
\end{aligned}
$$

We can simplify the inner sum further. Let $\boldsymbol{e} = b\lfloor \boldsymbol{a}/b \rfloor$, i.e. the ith component of $\boldsymbol{e}$ is given by $e_i = b\lfloor a_i/b \rfloor$. Further let $\boldsymbol{d} = \boldsymbol{a} - \boldsymbol{e}$, i.e. the ith component of $\boldsymbol{d}$ is given by $d_i = a_i - e_i$. Then we have

$$
\begin{aligned}
&\sum_{u \subseteq K} (-1)^{|u|} b^{-|u|} c_{\boldsymbol{l}-(\mathbf{1}_u,\mathbf{0}),(\lfloor \boldsymbol{a}_u/b \rfloor, \boldsymbol{a}_{\{1,\ldots,ds\}\setminus u})} \\
&\quad = \sum_{u \subseteq K} (-1)^{|u|} b^{-|u|} \sum_{\boldsymbol{k}_u \in A_{\mathbf{1}_u}} c_{\boldsymbol{l},\boldsymbol{e}+(\boldsymbol{k}_u,\boldsymbol{d}_{\{1,\ldots,ds\}\setminus u})} \\
&\quad = \sum_{u \subseteq K} (-1)^{|u|} b^{-|u|} b^{-ds+|u|} \sum_{\boldsymbol{k} \in A_{\mathbf{1}}} c_{\boldsymbol{l},\boldsymbol{e}+(\boldsymbol{k}_u,\boldsymbol{d}_{\{1,\ldots,ds\}\setminus u})} \\
&\quad = b^{-ds} \sum_{\boldsymbol{k} \in A_{\mathbf{1}}} \sum_{u \subseteq K} (-1)^{|u|} c_{\boldsymbol{l},\boldsymbol{a}+(\boldsymbol{k}_u-\boldsymbol{d}_u,\mathbf{0}_{\{1,\ldots,ds\}\setminus u})} \\
&\quad = b^{-ds} \sum_{\boldsymbol{k} \in A_{\mathbf{1}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\mathbf{1})b^{-\boldsymbol{l}}))} \sum_{u \subseteq K} (-1)^{|u|} f(\boldsymbol{t} + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u - \boldsymbol{d}_u, \mathbf{0}_{\{1,\ldots,ds\}\setminus u}))) \,\mathrm{d}\boldsymbol{t},
\end{aligned}
$$

where we extend the digit interlacing function $\mathscr{D}_d$ to negative values by using digits in $\{1 - b, \ldots, 0\}$ in case a component is negative. To shorten the notation we set

$$
\delta_{\boldsymbol{k}}(\boldsymbol{t}) = \sum_{u \subseteq K} (-1)^{|u|} f(\boldsymbol{t} + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u - \boldsymbol{d}_u, \mathbf{0}_{\{1,\ldots,\alpha s\}\setminus u}))).
$$

Therefore

$$
\begin{aligned}
\sigma_{d,\boldsymbol{l},s}^2(f) &\leq b^{|\boldsymbol{l}|_1-2ds} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} \sum_{\boldsymbol{k} \in A_{\mathbf{1}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\mathbf{1})b^{-\boldsymbol{l}}))} |\delta_{\boldsymbol{k}}(\boldsymbol{t})| \,\mathrm{d}\boldsymbol{t} \\
&\quad \times \sum_{\boldsymbol{k}' \in A_{\mathbf{1}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\mathbf{1})b^{-\boldsymbol{l}}))} |\delta_{\boldsymbol{k}'}(\boldsymbol{t})| \,\mathrm{d}\boldsymbol{t} \\
&= b^{|\boldsymbol{l}|_1-2ds} \sum_{\boldsymbol{k},\boldsymbol{k}' \in A_{\mathbf{1}}} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\mathbf{1})b^{-\boldsymbol{l}}))} |\delta_{\boldsymbol{k}}(\boldsymbol{t})| \,\mathrm{d}\boldsymbol{t} \\
&\quad \times \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\mathbf{1})b^{-\boldsymbol{l}}))} |\delta_{\boldsymbol{k}'}(\boldsymbol{t})| \,\mathrm{d}\boldsymbol{t}.
\end{aligned}
$$

Using Cauchy-Schwarz' inequality we have

$$\begin{aligned}
&\int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} |\delta_{\boldsymbol{k}}(\boldsymbol{t})| \,\mathrm{d}\boldsymbol{t} \\
&\le \left(\int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} 1 \,\mathrm{d}\boldsymbol{t}\right)^{1/2} \left(\int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} |\delta_{\boldsymbol{k}}(\boldsymbol{t})|^2 \,\mathrm{d}\boldsymbol{t}\right)^{1/2} \\
&= b^{-|\boldsymbol{l}|_1/2} \left(\int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} |\delta_{\boldsymbol{k}}(\boldsymbol{t})|^2 \,\mathrm{d}\boldsymbol{t}\right)^{1/2}.
\end{aligned}$$

Let $B_{\boldsymbol{a},\boldsymbol{k}} = \left(\int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} |\delta_{\boldsymbol{k}}(\boldsymbol{t})|^2 \,\mathrm{d}\boldsymbol{t}\right)^{1/2}$. Then we have

$$\begin{aligned}
\sigma^2_{d,\boldsymbol{l},s}(f) &\le b^{-2ds} \sum_{\boldsymbol{k},\boldsymbol{k}' \in A_{\boldsymbol{1}}} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} B_{\boldsymbol{a},\boldsymbol{k}} B_{\boldsymbol{a},\boldsymbol{k}'} \\
&\le \max_{\boldsymbol{k},\boldsymbol{k}' \in A_{\boldsymbol{1}}} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} B_{\boldsymbol{a},\boldsymbol{k}} B_{\boldsymbol{a},\boldsymbol{k}'} \\
&= \max_{\boldsymbol{k} \in A_{\boldsymbol{1}}} \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} B^2_{\boldsymbol{a},\boldsymbol{k}},
\end{aligned}$$

where the last inequality follows as the Cauchy-Schwarz inequality is an equality for two vectors which are linearly dependent. Let $\boldsymbol{k}^*$ be the value of $\boldsymbol{k} \in A_{\boldsymbol{1}}$ for which the sum $\sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} B^2_{\boldsymbol{a},\boldsymbol{k}}$ takes on its maximum. Hence

$$\sigma^2_{d,\boldsymbol{l},s}(f) \le \sum_{\boldsymbol{a} \in A_{\boldsymbol{l}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} |\delta_{\boldsymbol{k}^*}(\boldsymbol{t})|^2 \,\mathrm{d}\boldsymbol{t}.$$

The following lemma relates the function $\delta_{\boldsymbol{k}}$ to the divided differences introduced above.

Lemma 11 *Let $\boldsymbol{l}$, $\boldsymbol{a}$, $\boldsymbol{e}$, $\boldsymbol{q}$, K and $K_1, \ldots, K_s$ be defined as above. For $\boldsymbol{t} \in \mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}))$ we have*

$$|\delta_{\boldsymbol{k}^*}(\boldsymbol{t})| \le 2^{s(d-\alpha)} \sup |\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}'; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_s) f|$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)$ with $\alpha_i = \min(|K_i|, \alpha)$, and the supremum is taken over all $\boldsymbol{t}' \in \mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}))$ and $\boldsymbol{z}_i = (z_{i,1}, \ldots, z_{i,\alpha_i})$ with $z_{i,r_i} = \tau_{i,r_i} b^{-d(l_i-1)-r_i}$ where $\tau_{i,r_i} \in \{1-b, \ldots, b-1\} \setminus \{0\}$ for $1 \le r_i \le |K_i|$ and $1 \le i \le s$ and such that all the points at which f is evaluated in $\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}'; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_s)$ are in $\mathscr{D}_{\boldsymbol{\alpha}}([\boldsymbol{q}b^{-\boldsymbol{l}+\boldsymbol{1}_K}, (\boldsymbol{q}+\boldsymbol{1})b^{-\boldsymbol{l}+\boldsymbol{1}_K}))$. Furthermore, we may assume that $|z_{i,1}| < |z_{i,2}| < \cdots < |z_{i,|K_i|}|$ for $1 \le i \le s$.

Proof. We show that $\delta_{\boldsymbol{k}^*}(\boldsymbol{t})$ can be written as divided differences. Since the divided difference operators are applied to each coordinate separately, it suffices to show the result for $s=1$. In this case we have

$$\delta_{k^*}(t) = \sum_{u \subseteq K} (-1)^{|u|} f(t + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u^* - \boldsymbol{d}_u, \boldsymbol{0}_{\{1,\ldots,d\}\setminus u}))),$$

where now $K = \{i \in \{1,\ldots,d\} : l_i > 0\}$.

Let $\boldsymbol{l} = (l_1,\ldots,l_d)$. Let $t = \frac{t_1}{b} + \frac{t_2}{b^2} + \cdots$, $\boldsymbol{a} = (a_1,\ldots,a_d)$ and $a_j = a_{j,l_j} + a_{j,l_j-1}b + \cdots + a_{j,1}b^{l_j-1}$. Then for $t \in \mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}))$ we have

$$t_{j+(l-1)d} = a_{j,l} \quad \text{for } 1 \le l \le l_j \text{ and } j \in K.$$

Further we have $d_j = a_{j,l_j}$ for $j \in K$. Let

$$I = \{j + (l-1)d : 1 \le l \le l_j, j \in K\}.$$

Then for $t \in \mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}))$ and $u \subseteq K$ we have

$$t + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u - \boldsymbol{d}_u, \boldsymbol{0})) = \sum_{j \in K} \sum_{l=1}^{l_j-1} \frac{a_{j,l}}{b^{j+(l-1)d}} + \sum_{j \in u} \frac{k_j}{{}^{j+(l_j-1)d}} + \sum_{j \in K\setminus u} \frac{a_{j,l_j}}{b^{j+(l_j-1)d}} + \sum_{j \in \mathbb{N}\setminus I} \frac{t_j}{b^j}.$$

For given $t \in \mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}}, (\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}))$ let

$$\tau_u = t + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u^* - \boldsymbol{d}_u, \boldsymbol{0}_{\{1,\ldots,d\}\setminus u})).$$

Let $\boldsymbol{k}^* = (k_1^*,\ldots,k_d^*)$ and

$$z_j = \frac{k_j^* - a_{j,l_j}}{b^{j+(l_j-1)d}} \quad \text{for } j \in K.$$

Notice that if $z_j = 0$, then $\delta_{\boldsymbol{k}^*}(t) = 0$ and hence we can exclude this case. Then for $v \subset u \subseteq K$ we have

$$\tau_u - \tau_v = \sum_{j \in u\setminus v} z_j.$$

Therefore

$$\begin{aligned}
\delta_{k^*}(t) &= \sum_{u \subseteq K} (-1)^{|u|} f(t + \mathscr{D}_d(b^{-\boldsymbol{l}}(\boldsymbol{k}_u^* - \boldsymbol{d}_u, \boldsymbol{0}_{\{1,\ldots,d\}\setminus u}))) \\
&= \sum_{u \subseteq K} (-1)^{|u|} f(\tau_u)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{u\subseteq K}(-1)^{|u|}f(\tau_\emptyset + (\tau_u - \tau_\emptyset)) \\
&= \sum_{u\subseteq K}(-1)^{|u|}f(t + \sum_{j\in u} z_j) \\
&= \Delta_{|K|}(t;\boldsymbol{z}')f,
\end{aligned}
$$

where $\boldsymbol{z}' = (z_j)_{j\in K}$.

Notice that the ordering of the elements in $\boldsymbol{z}'$ does not change the value of $\Delta_{|K|}(t;\boldsymbol{z}')$. Hence assume that the elements in $\boldsymbol{z}'$ are ordered such that $z'_1 > z'_2 > \cdots > z'_{|K|}$. For the case where $|K| > \alpha$ we obtain from the definition of the divided differences that

$$\Delta_{|K|}(t;\boldsymbol{z}') = \sum_{u\subseteq\{|K|+1,\ldots,\alpha\}} (-1)^{|u|}\Delta_\alpha\left(t + \sum_{j\in u} z'_j; (z'_1,\ldots,z'_\alpha)\right).$$

By taking the triangular inequality and the supremum over all t' in $\{t + \sum_{j\in u} z'_j : u \subseteq \{|K|+1,\ldots,\alpha\}\}$, we obtain

$$\Delta_{|K|}(t;\boldsymbol{z}') \leq 2^{\alpha-|K|}\sup_{t'}|\Delta_\alpha(t';(z'_1,\ldots,z'_\alpha))|.$$

Consider now the general case $s \geq 1$ and $K = \{i \in \{1,\ldots,ds\} : l_i > 0\}$. Let $K_i = K \cap \{(i-1)d+1,\ldots,(i-1)d+d\}$ and $\alpha'_i = |K_i|$ for $1 \leq i \leq s$. Let $\boldsymbol{\alpha}' = (\alpha'_1,\ldots,\alpha'_s)$. Let

$$z_j = \frac{k^*_j - a_{j,l_j}}{b^{j-(i-1)d+(l_j-1)d}} \quad \text{for } j \in K_i \text{ and } 1 \leq i \leq s.$$

and $\boldsymbol{z}'_i = (z_j)_{j\in K_i}$ for $1 \leq i \leq s$. Then we obtain

$$\delta_{\boldsymbol{k}^*}(\boldsymbol{t}) = \Delta_{\boldsymbol{\alpha}'}(\boldsymbol{t};\boldsymbol{z}'_1,\ldots,\boldsymbol{z}'_s)f.$$

Define now $\alpha_i = \min(\alpha,\alpha'_i)$ for $1 \leq i \leq s$ and $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_s)$. Notice that $\alpha'_i \leq d$ and therefore

$$\sum_{i=1}^{s}(\alpha'_i - \alpha_i) \leq s(d-\alpha).$$

Notice that $\Delta_{\alpha'_i}$ can be expressed as a sum an alternating sum of $2^{\alpha'_i-\alpha_i}$ summands Δ_{α_i}.

By taking the triangular inequality we therefore obtain

$$|\delta_{\boldsymbol{k}^*}(\boldsymbol{t})| = |\Delta_{\boldsymbol{\alpha}'}(\boldsymbol{t};\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)f| \leq 2^{s(d-\alpha)}\sup|\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}';\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)|,$$

where the supremum is taken over all admissible choices of $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s$ and $\boldsymbol{t}'$. $\square$

Hence

$$\sigma_{d,\boldsymbol{l},s}^2(f) \le 2^{s(d-\alpha)} \sum_{\boldsymbol{a}\in A_{\boldsymbol{l}}} \int_{\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])} \sup |\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}';\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)f|^2 \,\mathrm{d}\boldsymbol{t},$$

where the supremum is over the same set as in Lemma 11. Therefore

$$\begin{aligned}
\sigma_{d,\boldsymbol{l},s}^2(f) &\le 2^{s(d-\alpha)} \sum_{\boldsymbol{a}\in A_{\boldsymbol{l}}} \mathrm{Vol}(\mathscr{D}_d([\boldsymbol{a}b^{-\boldsymbol{l}},(\boldsymbol{a}+\boldsymbol{1})b^{-\boldsymbol{l}}])) \sup |\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}';\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)f|^2 \\
&\le 2^{s(d-\alpha)} \sum_{\boldsymbol{q}\in A_{\boldsymbol{l}-\boldsymbol{1}_K}} \mathrm{Vol}(\mathscr{D}_d([\boldsymbol{q}b^{-\boldsymbol{l}+\boldsymbol{1}_K},(\boldsymbol{q}+\boldsymbol{1})b^{-\boldsymbol{l}+\boldsymbol{1}_K}])) \sup |\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t}';\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)f|^2.
\end{aligned}$$

Let $\gamma_j' = (b-1)b^{-j+(i-1)d-(l_j-1)d}$ for $j \in K_i$ and $1 \le i \le s$. Let $\gamma_{i,1} < \gamma_{i,2} < \cdots < \gamma_{i,\alpha_i}$ for $1 \le i \le s$ be such that $\{\gamma_{i,1},\ldots,\gamma_{i,\alpha_i}\} = \{\gamma_j : j \in K_i\}$, that is, $\{\gamma_{i,j} : 1 \le j \le \alpha_i\}$ is just a reordering of the elements of the set $\{\gamma_j : j \in K_i\}$. Set $\gamma(\boldsymbol{l}) = \prod_{i=1}^s \prod_{j=1}^{\alpha_i} \gamma_{i,j}$. Then

$$\begin{aligned}
\sigma_{\alpha,\boldsymbol{l},s}^2(f) &\le 2^{s(d-\alpha)}\gamma^2(\boldsymbol{l}) \sum_{\boldsymbol{e}\in A_{\boldsymbol{l}-\boldsymbol{1}_K}} \mathrm{Vol}(\mathscr{D}_\alpha([\boldsymbol{q}b^{-\boldsymbol{l}+\boldsymbol{1}_K},(\boldsymbol{q}+\boldsymbol{1})b^{-\boldsymbol{l}+\boldsymbol{1}_K}])) \sup \frac{|\Delta_{\boldsymbol{\alpha}}(\boldsymbol{t};\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s)f|^2}{\prod_{i\in K}|z_i|^2} \\
&\le 2^{s(d-\alpha)}\gamma^2(\boldsymbol{l}) V_\alpha^2(f),
\end{aligned}$$

where the supremum is over all admissible $\boldsymbol{t}$ and $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_s$ as described in the lemma.

Consider now the case where $K_i = \emptyset$ for some $1 \le i \le s$. Let $R = \{i \in \{1,\ldots,s\} : K_i = \emptyset\}$. Then the result follows by replacing f with the function $\int_{[0,1]^{|R|}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_R$ in the proof above.

Let now $d < \alpha$. Then $V_d(f) \le V_\alpha(f)$, and hence the result follows by using the proof above with $d = \alpha$. This completes the proof.

References

[1] N.S. Bakhvalov, On approximate computation of integrals. Vestnik MGV, Ser. Mat. Mekh. Astron. Fiz. Khim, 4, 3–18, 1959.

[2] J. Baldeaux, J. Dick, and F. Pillichshammer, A characterization of generalized nets using Weyl sums and its applications. Unif. Distr. Theory, 5, 133155, 2010.

[3] H. E. Chrestenson. A class of generalized Walsh functions. *Pacific J. Math.,* 5:17–31, 1955.

[4] J. Dick, Explicit constructions of quasi-Monte Carlo rules for the numerical integration of high dimensional periodic functions. SIAM J. Numer. Anal., 45, 2141–2176, 2007.

[5] J. Dick, Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order. SIAM J. Numer. Anal., 46, 1519–1553, 2008.

[6] J. Dick, On quasi-Monte Carlo rules achieving higher order convergence. In: Proceedings of the MCQMC08 conference, Montreal, Canada, P. LEcuyer and A. Owen (eds.), pp. 7396, 2009.

[7] J. Dick and F. Pillichshammer, *Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo integration.* Cambridge University Press, Cambridge, 2010.

[8] P. L'Ecuyer and C. Lemieux, A survey of randomized quasi-Monte Carlo. In: *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications.* M. Dror, P. L'Ecuyer, and F. Szidarovszki (eds.), 419–474. Kluwer Academic, New York, 2002.

[9] S. Heinrich, Random approximation in numerical analysis. In: *Functional Analysis*, K. D. Bierstedt et al. (eds.), 123-171. Dekker, New York, 1994.

[10] F. J. Hickernell, The mean square discrepancy of randomized nets, ACM Trans. Modeling Comput. Simul. 6, 274 296, 1996.

[11] J. Matoušek, On the L_2 discrepancy for anchored boxes. J. Complexity, 14, 527–556, 1998.

[12] J. Matoušek, *Geometric Discrepancy: An Illustrated Guide.* Springer, Heidelberg, 1998.

[13] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia, 1992.

[14] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, Springer Verlag, 1988.

[15] A.B. Owen, Randomly permuted (t, m, s)-nets and (t, s)-sequences. In: H. Niederreiter and J.-S. Shiue (eds.), *Monte Carlo and quasi-Monte Carlo methods in scientific computing* (Las Vegas, NV, 1994), 299–317, Lecture Notes in Statist., 106, Springer, New York, 1995.

[16] A. B. Owen, Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal., 34, 1884–1910, 1997.

[17] A. B. Owen, Scrambled net variance for integrals of smooth functions. Ann. Statist., 25, 1541–1562, 1997.

[18] A. B. Owen, Variance with alternative scramblings of digital nets. ACM Trans. Model. Comp. Simul., 13, 363–378, 2003.

[19] A. B. Owen, Local antithetic sampling with scrambled nets. Ann. Statist., 36, 2319–2343, 2008.

[20] S. Tezuka and H. Faure, I-binomial scrambling of digital nets and sequences. J. Complexity, 19, 744-757, 2003.

[21] J. L. Walsh. A Closed Set of Normal Orthogonal Functions. *Amer. J. Math.*, 45:5–24, 1923.

[22] R.X. Yue and F.J. Hickernell, The discrepancy and gain coefcients of scrambled digital nets, J. Complexity, 18, 135151, 2002.

Address:
Josef Dick
School of Mathematics and Statistics
The University of New South Wales
Sydney, 2052 NSW, Australia
Email: josef.dick@unsw.edu.au
Web: http://profiles.unsw.edu.au/maths/jdick1
Blog: http://quasirandomideas.wordpress.com/